輪読ゼミ・チョウ シホウ・12M42340

1 SVM の最適化

1.1 問題点

- 大きいデータサイズでは、汎用 QP ソルバーでは計算困難。
- ヘッセ行列は密（dense）と悪条件（ill-conditioned）。

1.2 SVM の主問題と双対問題

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \\ \text{subject to} \quad & y_i(w^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0\,, 1 \leq i \leq N. \end{aligned} \tag{1}$$

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{N}\alpha_i \\ \text{subject to} \quad & \sum_{i=1}^{N} y_i a_i = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned} \tag{2}$$

1.3 Karush-Kuhn-Tucker Conditions

$$\nabla f(x^*) + \sum_{i=1}^{m}\mu_i \nabla g_i(x^*) + \sum_{j=1}^{l}\lambda_j \nabla h_j(x^*) = 0$$

$$\text{主不等式制約：}\quad g_i(x^*) \leq 0,\ \text{for all } i = 1, \ldots, m$$

$$\text{主等式制約：}\quad h_j(x^*) = 0,\ \text{for all } j = 1, \ldots, l$$

$$\text{双対不等式制約：}\quad \mu_i \geq 0,\ \text{for all } i = 1, \ldots, m$$

$$\text{相補性条件：}\quad \mu_i g_i(x^*) = 0, \text{for all } i = 1, \ldots, m.$$

1.4 SVM の KKT 条件

$$L(w, b, \xi; \alpha, \beta) = \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i(1 - \xi_i - y_i(w^T x_i + b)) + \sum_{i=1}^{N}\beta_i(-\xi_i)$$

$$(1 - \xi_i - y_i(w^T x_i + b)) \leq 0, -\xi_i \leq 0 \tag{3}$$

$$\alpha_i \geq 0, \beta_i \geq 0 \tag{4}$$

$$\alpha_i(1-\xi_i-y_i(w^Tx_i+b))=0, \beta_i\xi_i=0 \tag{5}$$

$$\nabla w = w-\sum_{i=1}^{N}\alpha_i y_i x_i=0, \nabla b=\sum_{i=1}^{N} y_i\alpha_i=0, \nabla\xi_i = C-\alpha_i-\beta_i=0 \tag{6}$$

1.5 Kernel Trick

- 入力データを高次元な空間に（非線形）射像する
- 特徴空間で線形分離できる。
- 計算複雑度の増大を抑えられる。

$$\text{例：}\quad K(x,z)=(x^Tz+c)^2=\sum_{i,j=1}^{N}(x_ix_j)(z_iz_j)+\sum_{i=1}^{N}(\sqrt{2}cx_i\sqrt{2}cz_i)+c^2=\phi(x)^T\phi(z)$$

$$x=\begin{pmatrix}x_1\\x_2\end{pmatrix}, \phi(x)=\begin{pmatrix}x_1x_1\\x_1x_2\\x_2x_1\\x_2x_2\\\sqrt{2}cx_1\\\sqrt{2}cx_2\\c\end{pmatrix}$$

1.6 SVM の最適解

$$u=w^Tx+b=(\sum_{i=1}^{N}\alpha_iy_ix_i)^Tx+b=\sum_{i=1}^{N}\alpha_iy_iK(x_i,x)+b \tag{7}$$

1.7 Support Vector

$$a_i=0 \Leftrightarrow y_iu_i\geq 1$$

$$0<a_i<C \Leftrightarrow y_iu_i=1$$

$$a_i=C \Leftrightarrow y_iu_i\leq 1$$

2 伝統的なアプローチ

2.1 Chunking

Vapnik による戦略が最も簡単なものであるが、実用性には制限があると思われる。

- Chunking, Vapnik 1982

2.2 Decomposition

QP のサイズが大きくなるのを防ぐために、問題の大きさをあらかじめ制限してしまう方法である。サポートベクターの候補となる変数をすべて QP の変数として扱わずにその一部のみを動かすことによって、問題のサイズを小規模に止めて反復を繰り返す。

- Osuna's method [Osuna et al., 1997]
- Sequential minimal optimization/SMO [Platt, 1998]
- LIBSVM [Fan et al., 2005]
- $\mathrm{SVM}^{\mathrm{light}}$ [Joachims, 1999]

2.3 Sequential minimal optimization

1. α に初期値を設定する。
2. 2 変数 α_i, α_j を選択する。
3. α_i, α_j に関する 2 変数部分最適化問題を解く。
4. α と勾配の更新。
5. KKT 条件を判定し、満たせば終了、そうでない場合には 1 へ。

3 その他の SVM 最適化へのアプローチ

- 内点法 [Ferris and Munson, 2002][Gertz and Wright 2003]
- 主問題で解く [$\mathrm{SVM}^{\mathrm{perf}}$ Joachims 2006][Shalev-Shwartz et al.]
- オンライン SVM [Buttou et al. 2005]